*International Journal of Engineering Issues*
Vol. 2015, no. 2, pp. 85-90
ISSN: 2458-651X


# Codons in Data Safe Transfer

M. Yamuna, A. Elakkiya
*School of Advanced Sciences, VIT University, Vellore.*
Email: *myamuna@vit.ac.in*.

***Abstract-*** Biotechnological methods can be used for cryptography. This paper presents an implementation of steganography using DNA molecules. We first encode a plaintext message into a DNA sequence using a randomly generated single-substitution key using the general DNA codon table. This sequence is then mixed and inserted into a background DNA.

***Keywords*****:** *DNA Sequence; Encryption; Decryption.*

## I. INTRODUCTION

Before the modern era, cryptography was concerned with message confidentiality (i.e., encryption) - conversion of messages from a comprehensible form into an incomprehensible one and back again at the other end, rendering it unreadable by interceptors without secret knowledge i.e., the key needed for decrypting. DNA cryptography is one of the rapid emerging technology which works on concepts of DNA computing. A new technique for securing data was introduced using the biological structure of DNA called DNA Computing. The technique later on extended by various researchers for encrypting and reducing the storage size of data that made the data transmission over the network faster and secured. [1] S. Jeevidha et.al analyzed an existing approach to the DNA computing method and DNA based cryptographic approach [2]. A. Atito proposed a novel algorithm to communicate data securely. The proposed technique is a composition of both encryption and data hiding using some properties of Deoxyribonucleic Acid (DNA) sequences [3]. K. Menaka proposed an algorithm using DNA sequences for data hiding is proposed and discussed for secure data transmission and reception [4]. D. Umalkar et.al., purposed the formula to write the message with the premise of complementary rules deoxyribonucleic acid sequence [5]. Kritika Gupta et.al., done a review on DNA based cryptographic techniques [6]. Eihab Bashier et.al proposed a technique of generating artificial DNA sequences using chaotic maps running on the domain of integer numbers [7]. M.Yamuna proposed a method for safe transfer of messages using periodic table [8]. Fatma E. Ibrahim et.al proposed a new data hiding algorithm based on DNA sequence [9]. Nirmalya Kar et.al proposed a new algorithm which has two rounds of encryption . Moreover paper contains a new method of key generation and sharing for better security aspects. The encryption scheme is designed by using the mechanism of DNA sequencing [10]. M.Yamuna et.al., developed a new binary periodic table and hence use this for encrypting details about any drug [11].

## II. PRELIMINARY NOTE

In this section we provide some basic results required for the proposed method.

### *A. DNA Codon Table*

The genetic code is the set of rules by which information encoded within genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells. Biological decoding is accomplished by the ribosome, which links amino acids in an order specified by mRNA, using transfer RNA (tRNA) molecules to

carry amino acids and to read the mRNA three nucleotides at a time. The genetic code is highly similar among all organisms, and can be expressed in a simple table with 64 entries [12]

We use the standard codon table in snapshot – 1 [13] for our encryption scheme.

| nonpolar | polar | basic | acidic | (stop codon) |
|---|---|---|---|---|

**Standard genetic code**

| 1st base | 2nd base | | | | | | | | 3rd base |
|---|---|---|---|---|---|---|---|---|---|
| | T | | C | | A | | G | | |
| T | TTT | (Phe/F) Phenylalanine | TCT | (Ser/S) Serine | TAT | (Tyr/Y) Tyrosine | TGT | (Cys/C) Cysteine | T |
| | TTC | | TCC | | TAC | | TGC | | C |
| | TTA | (Leu/L) Leucine | TCA | | TAA | Stop (*Ochre*) | TGA | Stop (*Opal*) | A |
| | TTG | | TCG | | TAG | Stop (*Amber*) | TGG | (Trp/W) Tryptophan | G |
| C | CTT | | CCT | (Pro/P) Proline | CAT | (His/H) Histidine | CGT | (Arg/R) Arginine | T |
| | CTC | | CCC | | CAC | | CGC | | C |
| | CTA | | CCA | | CAA | (Gln/Q) Glutamine | CGA | | A |
| | CTG | | CCG | | CAG | | CGG | | G |
| A | ATT | (Ile/I) Isoleucine | ACT | (Thr/T) Threonine | AAT | (Asn/N) Asparagine | AGT | (Ser/S) Serine | T |
| | ATC | | ACC | | AAC | | AGC | | C |
| | ATA | | ACA | | AAA | (Lys/K) Lysine | AGA | (Arg/R) Arginine | A |
| | ATG[A] | (Met/M) Methionine | ACG | | AAG | | AGG | | G |
| G | GTT | (Val/V) Valine | GCT | (Ala/A) Alanine | GAT | (Asp/D) Aspartic acid | GGT | (Gly/G) Glycine | T |
| | GTC | | GCC | | GAC | | GGC | | C |
| | GTA | | GCA | | GAA | (Glu/E) Glutamic acid | GGA | | A |
| | GTG | | GCG | | GAG | | GGG | | G |

**Snapshot – 1**

### *B. Insertion Method*

In [ 14 ], H.J. Shiu et al introduced the insertion method, which is wide use now. Snapshot [ 14 ] of insertion method is given below

**Method 1: the Insertion Method**

To simplify the discussion, the most basic version is outlined and a simple example is given. The more complicated version of the method will be presented after the basic one is explained. All of the methods use a reference sequence, *S*. Suppose the secret message *M* is 01001100. Let *S* be *ACGGTTCCAATGC*. The method works as follows:

**Step 1.** Code *S* into a binary sequence by using the binary coding rule. Thus the sequence *S* will now become 00011010111101010000111001.

**Step 2.** Divide *S* into segments, whereby each segment contains *k* bits. Suppose *k* is 3. Then there are the following segments: 000, 110, 101, 111, 010, 100, 001, 110, 01.

**Step 3.** Insert bits from *M*, one at a time, into the beginning of segments of *S*. The result is as follows: 0000, 1110, 0101, 0111, 1010, 1100, 0001, 0110, 01. Those segments without any secret message inserted should be ignored. Thus, there are the following segments: 0000, 1110, 0101, 0111, 1010, 1100, 0001, 0110. Concatenating the above segments results in the following binary sequence: 00001110010101111010110000010110.

**Step 4.** Use the inverse function of the binary coding rule to produce the following faked DNA sequence: *S′* = *AATGCCCTGGTAACCG*. As the reader can see, this sequence is quite different from *S*.

**Step 5.** Send the above sequence *S′* to the receiver.

**Snapshot – 2**

## III. METHODS

### A. Proposed Encryption Scheme

The sequence of the bases A, T, G, C in DNA determines our unique genetic code. This also provides instructions for producing molecules. There are 20 different types of amino acids which are the building blocks of proteins. Different proteins are made up of combination of amino acids. The 20 amino acids are in general represented using 64 codons. DNA sequences represent the genetic privacy in medical field. Numerous data are available in public domain. We use the availability of these sequences for encrypting normal data. We use the regular codon table for constructing a new table which can be used for our encryption method.

### B. Construction of Modified DNA Codon Table

We use the general codon table. There are 64 entries in the table. We randomly pick any 27 combinations ( one for each alphabet and one for blank space ) and assign them in the table. Table – 1 represents a randomly constructed table.

| | **T** | | **C** | | **A** | | **G** | | |
|---|---|---|---|---|---|---|---|---|---|
| **T** | TTT<br>TTC | A<br>B | TCT<br>TCC | E<br>F | TAT<br>TAC | I<br>J | TGT<br>TGC | M<br>N | **T**<br>**C** |
| | TTA<br>TTG | C<br>D | TCA<br>TCG | G<br>H | TAA<br>TAG | K<br>L | TGA<br>TGG | O<br>P | **A**<br>**G** |
| **C** | CTT<br>CTC | Q<br>R | CCT<br>CCC | T<br>U | CAT<br>CAC | X<br>Y | CGT<br>CGC | | **T**<br>**C** |
| | CTA<br>CTG | S | CCA<br>CCG | V<br>W | CAA<br>CAG | Z<br>Space | CGA<br>CGG | | **A**<br>**G** |
| **A** | ATT<br>ATC<br>ATA | | ACT<br>ACC<br>ACA<br>ACG | | AAT<br>AAC | | AGT<br>AGC | | **T**<br>**C** |
| | | | | | AAA<br>AAG | | AGA<br>AGG | | **A**<br>**G** |
| | ATG | | | | | | | | |
| **G** | GTT<br>GTC<br>GTA<br>GTG | | GCT<br>GCC<br>GCA<br>GCG | | GAT<br>GAC | | GGT<br>GGC<br>GGA<br>GGG | | **T**<br>**C**<br>**A**<br>**G** |
| | | | | | GAA<br>GAG | | | | |

Table – 1

As seen in the table each alphabet is now represented by a codon. In Table – 1 A is represented as TTT, B as TTC and so on.

In regular insertion method, we choose a random sequence S. This string S is split into segments of size k. The message M to be encrypted is inserted between these k bits. Since insertion is traditionally used, once the size of the segment is determined, decryption becomes relatively easy. To overcome this, we adopt a random insertion rather then inserting the message M into all the segments.

### C. Modified Insertion Method

We use the insertion method slightly modified as random insertion, where after dividing the segment into segments of length k = 3 we can insert any bit of length 3 randomly at any place.
For example if the DNA sequence to be inserted is GATAAGGAC.

- We pick any random sequence of codons GGTAGCAGTAATACTATTATCGTT.
- We split them into segments of length k = 3 to generate GGT AGC AGT AAT ACT ATT ATC GTT.
- We then insert the codons in the original sequence randomly. The codons in the original sequence is GAT AAG GAC.
- This randomly inserted generates GGT AGC GAT AGT AAT AAG ACT ATT ATC GAC GTT.

### D. Encryption Algorithm

There are 64 entries in the codon table. Hence we can use this method for encrypting any message containing a maximum of 64 distinct characters. Table – 1 presents a sample table for normal message encryption using alphabets
Let M be the message to be encrypted.

**Step 1** Construct a modified codon table as in Table – 1.
**Step 2** Assign the codons to the characters in the message M using Table – 1.
**Step 3** Choose a random DNA sequence S of any length not containing the codons used in Table – 1.
**Step 4** Insert the codons in M randomly into the sequence S to generate a sequence D.
**Step 5** Send the message D to the receiver.

### E. Decryption Algorithm

For decrypting the message we reverse the procedure.
Suppose the received message is S. Split the sequence S into codons. Filter the codons used in Table – 1 separately. Convert these codons into message using Table – 1.

## IV. DISCUSSION

### A. Illustration of the Encryption Algorithm

Let M = ALGEBRA be the message to be encrypted

| ***A*** | ***L*** | ***G*** | ***E*** | ***B*** | ***R*** | ***A*** |
|---|---|---|---|---|---|---|
| \| | \| | \| | \| | \| | \| | \| |
| ***TTT*** | ***TAG*** | ***TCA*** | ***TCT*** | ***TTC*** | ***CTC*** | ***TTT*** |

M = **TTT TAG TCA TCT TTC CTC TTT**

Let ATTATCACTACCCGGAGTAGC be the random sequence. Split this sequence into segments of length k = 3, to generate ATT ATC ACT ACC CGG AGT AGC. Then we insert M into this sequence randomly. The resulting sequence is

D: **ATT TTT ATC ACT TAG ACC TCA CGG TCT AGT TTC AGC CTC ATT TTT**

Send the sequence
**D: ATTTTTATCACTTAGACCTCACGGTCTAGTTTAGCCTCATTTTT** to the receiver.

### *B. Illustration of the Decryption Algorithm*

Suppose the received message S: ATTTCACGGTGAACCTGAATTAGCTTG ACTCAGCGG TAGCCCAC TT ATCTAA.

Split the sequence into codons. Filter the codons used in Table – 1 separately.

**D = ATTTCA CGGTGAACCTGAATTAGCTTGACTCAGCGGTAGCCCACTTTAAT CTAA.**

**S = TCA TGA TGA TTG CAG TAG CCC TTA TAA.**

Convert these codons into message using Table – 1.

| TCA | TGA | TGA | TTG | CAG | TAG | CCC | TTA | TAA |
|---|---|---|---|---|---|---|---|---|
| \| | \| | \| | \| | \| | \| | \| | \| | \| |
| G | O | O | D | Space | L | U | C | K |

The message is decrypted as GOOD LUCK.

Since insertion method is modified and adopted. Once the modified codon table is decided, the proposed method can be programmed easily in any of the common programming languages.

## V. CONCLUSION

DNA cryptography is encrypting the message in terms of DNA sequences. This can be done using several DNA technologies with the biochemical methods. This method can also be interpreted with other schemes in order to apply this technology in various fields. In this paper we use codons and DNA sequence to encrypt the message.

- For normal text encryption we use only 27 characters. These can be picked from 64 codons in $64C_{27}$ ways.
- The remaining 64 – 27 = 37 codons can be used to generate the DNA sequence for insertion technique.
- The message DNA is inserted into another DNA sequence. So it is difficult to determine which part is the encrypted DNA
- Even if the encrypted DNA is filtered it is not possible to decrypt the message unless it is known how the table is constructed.

The original insertion method is modified to random one. Also numerous DNA sequences are available in public domain. It is difficult to find the difference between the original and the encrypted one. So we conclude that the proposed method is safe for message transfer.

## REFERENCES

[1] http://securityaffairs.co/wordpress/33879/security/dna-cryptography.html.

[2] JEEVIDHA, S., BASHA, Dr MS Saleem, et DHAVACHELVAN, Dr P. Analysis on DNA based cryptography to secure data transmission. *International Journal of Computer Applications (0975–8887)*, 2011, vol. 29, no 8.

[3] ATITO, Ahmed, KHALIFA, A., et RIDA, S. Z. Dna-based data encryption and hiding using playfair and insertion techniques. *Journal of Communications & Computer Engineering*, 2012, vol. 2, no 3, p. 44-49.

[4] MENAKA, K., Message Encryption Using DNA Sequences, *World Congress on Computing and Communication Technologies,* 2014, p 182 – 18.

[5] Ms. Amruta D. Umalkar, Pritish A. Tijare, Data Encryption Using DNA Sequences Based On Complementary Rules – A Review, *International Journal of Engineering Research and General Science,* 2014, vol. 2, no. 6, p. 345-349.

[6] KRITIKA, Gupta, SHAILENDRA, Singh, DNA Based Cryptographic Techniques: A Review, *International Journal of Advanced Research in Computer Science and Software Engineering,* 2013, vol. 3, Issue 3, p. 607-610.

[7] EIHAB Bashier, GHAIDAA Ahmed, HUSSAM-ALDEEN Othman. Hiding Secret Messages using Artificial DNA Sequences Generated by Integer Chaotic Maps, *International Journal of Computer Applications*, 2013, vol. 70 , no. 15, p. 0975- 8887.

[8] YAMUNA, M. Data Encryption using Periodic Table, *The International Journal of Computer Science & Applications ,* 2014, vol. .3, no. 01, 2014.

[9] http://webcache.googleusercontent.com/search?q=cache:xifJkQUXwJkJ:bu.edu.eg/portal/uploads/Computers%25 20and%2520Informatics/Computer%2520System%2520/4974/publications/Mahmoud%2520Ibrahim%2520Mahmoud%2520Mossa_Enhancing%2520the%2520Security%2520of%2520Data%2520Hiding%2520Using%2520Double%2520DNA%2520Sequences.pdf+&cd=1&hl=en&ct=clnk&gl=in.

[10] NIRMALYA Kar, ATANU Majumder, ASHIM Saha, SUMAN Deb, Data Security And Cryptography Based On DNA Sequencing, *International Journal of Information Technology & Computer Science,* 2013 vol. 10, no 3, p. 24 – 32.

[11] YAMUNA, M., KARTHIKA, K. Periodic table as a Binary table for Drug Encryption, *International Journal of Drug Development and Research*, Vol. 6 , Issue 2, pp 52 – 56, 2014.

[12] http://en.wikipedia.org/wiki/Genetic_code.

[13] https://en.wikipedia.org/wiki/DNA_codon_table.

[14] SHIU H.J, NG K.L, FANG J.F, LEE R.C.T, HUANG C.H. Data hiding methods based upon DNA sequences, *Information Sciences*, pp 2196 – 2208, 2010.